11: Encoder
12, 22: Grammar rule
13: Compressor
21: Decompressor
23: Decoder
201: XML document
202: Structure part
203: Encoded data string
204: Contents
205: Compressed XML document

4. Claims:

Claim 1

A data compression apparatus for encoding data and for compressing the encoded data comprising:

a grammar storage unit for storing grammar rules for a tree local language in which data are represented by a labeled tree structure;

an encoder for reading a document written in said tree local language, for dividing said document into a structure part and contents, and for encoding said structure part using said grammar rules stored in said grammar storage unit; and

a compressor for compressing said contents of said document extracted by said encoder and for encoding the compressed contents.

Claim 2

The data compression apparatus according to claim 1, wherein said encoder includes:

a divider for dividing a target document into a structure part and contents;

an automata constructor, for constructing pushdown automata that correspond to said grammar rules; and

an encoded data generator, for employing said pushdown automata that are constructed by said automata constructor to perform syntax analyzation of said structure part of said document that is obtained by said divider, and for generating an encoded data string for said structure part.

Claim 3

The data compression apparatus according to claim 2, wherein said encoded data generator of said encoder assigns symbols to choices resident in said pushdown automata that are constructed by said automata constructor; and wherein said encoded data generator employs said pushdown automata to analyze said structure part of said document written in said tree local language, and outputs, at the location of the selected choices, said symbols that are assigned for said choices, so that the encoded data string for said structure part is generated.

Claim 4

The data compression apparatus according to claim 1, wherein said compressor performs compression and encoding not only for said contents of said document written in said tree local language, but also for said structure part of said document that is obtained by said encoder.

Claim 5

A data communication system comprising:

a transmission source data processing apparatus for transmitting data across a communication network; and

a transmission destination data processing apparatus for receiving said data transmitted across said communication network by said transmission source data processing apparatus,

said transmission source data processing apparatus including

a first grammar storage unit for storing grammar rules for a tree local language in which data are represented by a labeled tree structure,

an encoder for reading a document written in said tree local language, for dividing said document into a structure part and contents, and for encoding said structure part using said grammar rules stored in said first grammar storage unit,

a compressor for compressing said contents of said document extracted by said encoder and for encoding the compressed contents, and

a transmitter for transmitting said structure part encoded by said encoder and said contents compressed and encoded by said compressor, and

said transmission destination data processing apparatus including

a receiver for receiving data from said transmission source data processing apparatus,

a second grammar storage unit for storing the same grammar rules as said grammar rules stored in said first grammar storage unit of said transmission source data processing apparatus,

a decompressor for employing a decompression method, which corresponds to the compression and encoding method used by said compressor of said transmission source data processing apparatus, to decompress data that are

received by said receiver and that correspond to said contents of said document, and

a decoder for employing said grammar rules stored in said second grammar rule storage unit to decode the data that are received by said receiver and that correspond to said structure part of said document.

Claim 6

A database system for storing and managing data in a storage unit comprising:

a grammar storage unit for storing grammar rules for a tree local language in which data are represented by a labeled tree structure;

an encoder for reading a document written in said tree local language, for dividing said document into a structure part and contents, and for encoding said structure part using said grammar rules stored in said grammar storage unit;

a compressor for compressing said contents of said document extracted by said encoder and for encoding the compressed contents; and

a storage unit for storing said structure part of said document encoded by said encoder, and for storing said contents of said document that are compressed and encoded by said compressor.

Claim 7

The database system according to claim 6, wherein said compressor performs compression and encoding not only for said contents of said document written in said tree local language, but also for said structure part of said document that is obtained by said encoder.

Claim 8

A data compression method for encoding data and for compressing the encoded data comprising the steps of:

reading a document written in a tree local language in which data are represented by a labeled tree structure, and dividing said document into a structure part and contents;

encoding said structure part using said grammar rules for said tree local language; and

compressing said contents of said document extracted by said encoder and encoding the compressed contents.

Claim 9

The data compression method according to claim 8, wherein said step of encoding said structure part of said document includes the steps of:

constructing pushdown automata that correspond to said grammar rules;

assigning symbols to choices resident in said pushdown automata;

employing said pushdown automata to analyze said structure part of said document in accordance with the depth-first searching, and to output, at the locations of said choices, said symbols that are assigned to said choices; and

outputting a symbol string that is obtained by employing said pushdown automata as encoded data strings of said structure part of said document that is written in said tree local language.

Claim 10

The data compression method according to claim 9, further

comprising:

a step, to be performed before said step of encoding said structure part of said document written in said tree local language and when an attribute belongs to a node of a target document in said tree local language, of changing said attribute to a child node of an element possessing said attribute so as to convert said grammar rules of said tree local language and said document into a tree structure that is to be processed by said pushdown automata.

Claim 11

The data compression method according to claim 8, further comprising: a step, to be performed after said step of encoding said structure part of said document, of employing another general-purpose compression and encoding method to further compress and encode said encoded structure part of said document.

Claim 12

A storage medium on which input means of a computer stores a computer-readable program, said program permitting said computer to perform:

a process for reading a document written in a tree local language in which data are represented by a labeled tree structure, and for dividing said document into a structure part and contents;

a process for encoding said structure part using said grammar rules of said tree local language; and

a process for compressing said contents of said document extracted by said encoder and for encoding the compressed contents.

Claim 13

A program transmission apparatus comprising:

storage means for storing a program that permits a computer to perform

a process for reading a document written in a tree local language in which data are represented by a labeled tree structure, and for dividing said document into a structure part and contents,

a process for encoding said structure part using said grammar rules for said tree local language, and

a process for compressing said contents of said document extracted by said encoder and for encoding the compressed contents; and

transmission means for reading said program from said storage means, and for transmitting said program.